Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements

Ali H. Omar,[1] Jae-Gwang Won,[2] David M. Winker,[1] Soon-Chang Yoon,[2] Oleg Dubovik,[3] and M. Patrick McCormick[4]

Received 6 April 2004; revised 7 July 2004; accepted 3 August 2004; published 1 March 2005.

[1] We use radiance measurements and inversions of the Aerosol Robotic Network (AERONET) (Dubovik and King, 2000; Holben et al., 1998; Holben et al., 2001) to classify global atmospheric aerosols using the complete archive of the AERONET data set as of December 2002 and dating back to 1993 for some sites. More than 143,000 records of AERONET solar radiance measurements, derived aerosol size distributions, and complex refractive indices are used to generate the optical properties of the aerosol at more than 250 sites worldwide. Each record is used in a clustering algorithm as an object, with 26 variables comprising both microphysical and optical properties to obtain six significant clusters. Using the mean values of the optical and microphysical properties together with the geographic locations, we identified these clusters as desert dust, biomass burning, urban industrial pollution, rural background, polluted marine, and dirty pollution. When the records in each cluster are subdivided by optical depth class, the trends of the class size distributions show that the extensive properties (mode amplitude and total volume) vary by optical depth, while the intensive properties (mean radius and standard deviation) are relatively constant. Seasonal variations of aerosol types are consistent with observed trends. In particular, the periods of intense biomass burning activity and desert dust generation can be discerned from the data and the results of the analyses. Sensitivity and uncertainty analyses show that the clustering algorithm is quite robust. When subsets of the data set are randomly created and the clustering algorithm applied, we found that more than 94% of the records retain their classification. Adding 10% random noise to the microphysical properties and propagating this error through the scattering calculations, followed by the clustering algorithm, results in a misclassification rate of less than 9% when compared with the noise-free data.

Citation: Omar, A. H., J.-G. Won, D. M. Winker, S.-C. Yoon, O. Dubovik, and M. P. McCormick (2005), Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, *J. Geophys. Res.*, *110*, D10S14, doi:10.1029/2004JD004874.

1. Introduction

[2] Global aerosol properties are important not only for use in radiative transfer models but also as inputs to the inversion algorithms of satellite-based measurements of aerosol optical properties. As we approach an era of unprecedented global coverage of aerosol profile measurements (Geoscience Laser Altimeter System (GLAS) and Cloud Aerosol Lidar and Infrared Pathfinder Spaceborne Observations (CALIPSO)) to augment the satellite measurements of aerosol optical depths by passive instruments (Moderate Resolution Imaging Spectrometer (MODIS), Multiangle Imaging Spectroradiometer (MISR), POLDER, Total Ozone Mapping Spectrometer (TOMS), and Stratospheric Aerosol and Gas Experiment (SAGE)), there is a need for well characterized aerosol properties. In its most recent report, the *Intergovernmental Panel on Climate Change* (*IPCC*) [2001] found that radiative forcing by aerosols is the most uncertain of all radiative forcing estimates. Reducing these uncertainties calls for expanded aerosol measurements and studies to characterize different types of aerosols and sources [*National Research Council* (*NRC*), 1996] (see also the National Aerosol-Climate Interactions Program (NACIP) White Paper; available at http://www-NACIP/ucsd.edu). Tropospheric aerosols are diverse and their properties depend on sources, emission rates, and removal mechanisms which can be highly variable. To understand and quantify aerosol effects, there have been numerous domestic and international campaigns to characterize aerosol physical and chemical properties and processes. These include the Aerosol Characterization Experiments (ACE-1, ACE-2, and ACE-Asia), the Tropospheric Aerosol

[1]NASA Langley Research Center, Atmospheric Sciences, Hampton, Virginia, USA.
[2]School of Earth and Environmental Sciences, Seoul National University, Seoul, Korea.
[3]Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland, USA.
[4]Center for Atmospheric Sciences, Hampton University, Hampton, Virginia, USA.

0148-0227/05/2004JD004874$09.00

Radiative Forcing Observational Experiment (TARFOX), the Smoke, Clouds, and Radiation-B (SCAR-B) experiment, and the Indian Ocean Experiment (INDOEX). ACE-1 took place south of Australia in November–December 1995 and measured properties of the natural aerosol in the remote marine boundary layer [*Bates et al.*, 1998]. ACE-2 took place in the north Atlantic ocean in June–July 1997, and focused on the radiative effects and processes controlling anthropogenic aerosols from Europe and desert dust from Africa as they were transported over the north Atlantic ocean [*Raes et al.*, 1999]. ACE-Asia took place during the spring of 2001 off the coast of China, Japan and Korea. This region includes many types of aerosols of widely varying composition and sizes derived from one of the largest aerosol source regions on Earth. TARFOX [*Russell et al.*, 1999], designed to measure and analyze aerosol properties and effects along the US eastern seaboard, took place on 10–31 July 1996. INDOEX [*Ramanathan et al.*, 2001] measured aerosol properties over the tropical Indian Ocean in 1998 and 1999. The large aerosol optical thicknesses ($\sim$0.5) and the prominent role of the carbonaceous aerosols in the extinction budget during most of INDOEX underline the need to develop long-term records of specific species.

[3] In many cases, mean properties sorted by location or type to represent aerosol characteristics are used in radiative transfer calculations [*Kiehl and Briegleb*, 1993; *Kiehl and Rodhe*, 1995; *Nemesure et al.*, 1995; *Penner et al.*, 1992] or inversion algorithms of satellite measurements [*Tanré et al.*, 1999]. Assigning a set of mean aerosol physical and chemical properties to a given location based on a long-term average has significant shortcomings. This is because at any given location, the aerosol type can be variable on timescales as short as a few hours [*Sheridan et al.*, 2001]. These variations result from transport of distinct air masses to a site and nonsystematic events such as fires, wind gusts, hurricanes, tornadoes, and land clearing and development activities. These variations lead to diverse aerosol characteristics at each site on timescales as short as a few hours and preclude the long term average of properties from being a good representation of the characteristics for a site or region. On the other hand, the frequency of occurrence of a given type of aerosol at a location is an indication of the likelihood of finding that type of aerosol at that location if an adequately large sample (years) is used to calculate this frequency. Aerosol optical measurements must therefore be made at short timescales (on the order of a few hours) to develop a large database which can be used to derive statistically significant correlations. The Aerosol Robotic Network (AERONET) [*Holben et al.*, 1998, 2001] measurements provide a database with a fine temporal resolution albeit for column rather than vertically resolved measurements. AERONET is an automatic robotic Sun and sky scanning measurement network that has grown rapidly to over 200 sites worldwide. AERONET uses multiangle radiance measurements to retrieve the discrete aerosol size distributions in 22 size bins ranging from 0.05 to 15 μm and the complex refractive index. The network has the important features of uniform data collection, calibration, and data processing procedures. This study uses the whole AERONET archive (up to December 2002) of measurements and inversions to develop a type-specific set of mean optical properties of aerosols. Cluster analysis is used for categorization of atmospheric aerosol types. Six significant types are suggested by the cluster analysis which we identify as: desert dust, biomass burning, polluted continental, clean continental, polluted marine aerosol, and dirty pollution. In this classification, clean continental refers to a lightly loaded soot-free pollution normally found in rural areas and is good approximation for background aerosol. Dirty pollution refers to pollution containing significant amounts of absorbing species.

10. Conclusion

[38] A global data set, AERONET, has been used to identify main clusters of aerosol types and to determine microphysical properties of aerosol groups. The clustering algorithm objectively groups all the 143,000+ records examined into six categories. Using the mean values of the optical and microphysical properties together with the geographic locations, we identified these categories as desert dust, biomass burning, urban industrial pollution, rural background, polluted marine, and dirty pollution and presented the mean properties of these aerosol models. The effects of aerosol particle nonsphericity on the cluster analysis are unknown but not negligible. We expect that these models will enhance the available database of the characteristics of aerosol types. Since the cluster analysis assigned a category to each record, it is possible to examine the frequency of occurrence of different types of aerosols at each station. The data showed periods of intense biomass burning activity and desert dust generation consistent with independent observations. The variation of the extensive and intensive size distribution properties within categories showed consistent trends. In particular, when each cluster was subdivided by optical depth class, the trends of the class size distributions show that the extensive properties (mode amplitude and total volume) vary by optical depth while the intensive properties (mean radius and standard deviation) are relatively constant. The uncertainty and sensitivity tests showed that the clustering algorithm is quite robust and reproduces more than 94% of the classification when the data is arbitrarily halved. When random errors of 10% are added to the microphysical properties and propagated through the optical properties and the clustering algorithm, the records are correctly classified at a rate of at least 91%.